

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
15 August 2002 (15.08.2002)

PCT

(10) International Publication Number  
WO 02/063601 A1

(51) International Patent Classification<sup>7</sup>: G09G 5/00, 5/08, G03B 21/00

(21) International Application Number: PCT/US02/03433

(22) International Filing Date: 5 February 2002 (05.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
09/777,778 5 February 2001 (05.02.2001) US

(71) Applicant: CANESTA, INC. [US/US]; 2833 Junction Avenue, Suite 200, San Jose, CA 95134 (US).

(72) Inventors: RAFIL, Abbas; 1546 Wisteria Court, Los Altos, CA 94024 (US). BAMJI, Cyrus; 1074 Ocaso Camino, Fremont, CA 94539 (US). SZE, Cheng-Feng; Apartment 90, 20200 Lucille Avenue, Cupertino, CA 95014 (US).

(74) Agents: KAUFMAN, Michael, A. et al.; Flehr Hohbach Test Albritton & Herbert LLP, Suite 3400, 4 Embarcadero Center, San Francisco, CA 94111-4187 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

**Published:**

- with international search report
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHOD AND SYSTEM TO PRESENT IMMERSION VIRTUAL SIMULATIONS USING THREE-DIMENSIONAL MEASUREMENT

(57) Abstract: A virtual simulation system generates an image of a virtual control on a display that may be a heads-up-display in a vehicle. The system uses three-dimensional range finding data to determine when a user is sufficiently close to the virtual control to "manipulate" the virtual control. The user "manipulation" is sensed non-haptically by the system, which causes the displayed control image to move in response to user manipulation. System output is coupled, linearly or otherwise, to an actual device having a parameter that is adjusted substantially in real-time by user-manipulation of the virtual image. System generated displays can be dynamic and change appearance when a user's hand is in close proximity displays can disappear until needed, or can include menus and icons to be selected by the user who points towards or touches the virtual images. System generate images can include representation of the user for use in a training or gaming system.



WO 02/063601 A1

5

**METHOD AND SYSTEM TO PRESENT IMMERSION VIRTUAL  
SIMULATIONS USING THREE-DIMENSIONAL MEASUREMENT**

10

RELATION TO PREVIOUSLY FILED APPLICATION

Priority is claimed from U.S. provisional patent application, serial number 60/180,473 filed 3 February 2000, and entitled "User Immersion in Computer Simulations and Applications Using 3-D Measurement, Abbas Rafii and Cyrus  
15 Bamji, applicants.

FIELD OF THE INVENTION

The present invention relates generally to so-called virtual simulation methods and systems, and more particularly to creating simulations using three-dimensionally acquired data so as to appear immerse the user in what is being  
20 simulated, and to permit the user to manipulate real objects by interacting with a virtual object.

BACKGROUND OF THE INVENTION

So-called virtual reality systems have been computer implemented to mimic a  
25 real or a hypothetical environment. In a computer game context, for example, a user or player may wear a glove or a body suit that contains sensors to detect movement, and may wear goggles that present a computer rendered view of a real or virtual environment. User movement can cause the viewed image to change, for example to zoom left or right as the user turns. In some applications,  
30 the imagery may be projected rather than viewed through goggles worn by the user. Typically rules of behavior or interaction among objects in the virtual imagery being viewed are defined and adhered to by the computer system that controls the simulation. U.S. patent no. 5,963,891 to Walker (1999) entitled "System for Tracking Body Movements in a Virtual Reality System" discloses a  
35 system in which the user must wear a data-gathering body suit. U.S. patent no. 5,337,758 to Moore (1994) entitled "Spine Motion Analyzer and Method"

discloses a sensor-type suit that can include sensory transducers and gyroscopes to relay back information as to the position of a user's body.

5 In training type applications, aircraft flight simulators may be implemented in which a pilot trainee (e.g., a user) views a computer-rendered three-dimensional representation of the environment while manipulating controls similar to those found on an actual aircraft. As the user manipulates the controls, the simulated aircraft appears to react, and the three-dimensional environment is made to change accordingly. The result is that the user interacts with the rendered  
10 objects in the viewed image.

But the necessity to provide and wear sensor-implemented body suits, gloves, helmets, or the necessity to wear goggles can add to the cost of a computer simulated system, and can be cumbersome to the user. Not only is freedom of  
15 motion restricted by such sensor-implemented devices, but is often necessary to provide such devices in a variety of sizes, e.g., large-sized gloves for adults, medium-sized gloves, small-sized gloves, etc. Further, only the one user wearing the body suit, glove, helmet, goggles can utilize the virtual system; onlookers for example see essentially nothing. An onlooker not wearing such  
20 sensor-laden garments cannot participate in the virtual world being presented and cannot manipulate virtual objects.

U.S. patent no. 5,168,531 to Sigel (1992 entitled "Real-time Recognition of Pointing Information From Video" discloses a luminosity-based two-dimensional  
25 information acquisition system. Sigel attempts to recognize the occurrence of a predefined object in an image by receiving image data that is convolved with a set of predefined functions, in an attempt to define occurrences of elementary features characteristic of the predefined object. But Sigel's reliance upon luminosity data requires a user's hand to exhibit good contrast against a  
30 background environment to prevent confusion with the recognition algorithm used.

Two-dimensional data acquisition systems such as disclosed by Korth in U.S.

patent no. 5,767,842 (1998) entitled "Method and Device for Optical Input of Commands or Data use video cameras to image the user's hand or body. In some applications the images can be combined with computer-generated images of a virtual background or environment. Techniques including edge and shape  
5 detection and tracking, object and user detection and tracking, color and gesture tracking, motion detection, brightness and hue detection are sometimes used to try to identify and track user action. In a game application, a user could actually see himself or herself throwing a basketball in a virtual basketball court, for example, or shooting a weapon towards a virtual target. Such systems are  
10 sometimes referred to as immersion systems.

But two-dimensional data acquisition systems only show user motion in two dimension, e.g., x-axis, y-axis but not also z-axis. Thus if the user in real life would use a back and forth motion to accomplish a task, e.g., to throw a ball, in  
15 two-dimensional systems the user must instead substitute a sideways motion, to accommodate the limitations of the data acquisition system. In a training application, if the user were to pick up a component, rotate the component and perhaps move the component backwards and forwards, the acquisition system would be highly challenged to capture all gestures and motions. Also, such  
20 systems do not provide depth information, and such data that is acquired is luminosity-based and is very subject to ambient light and contrast conditions. An object moved against a background of similar color and contrast would be very difficult to track using such prior art two-dimensional acquisition systems. Further, such prior art systems can be expensive to implement in that  
25 considerable computational power is required to attempt to resolve the acquired images.

Prior art systems that attempt to acquire three-dimensional data using multiple two-dimensional video cameras similarly require substantial computing power,  
30 good ambient lighting conditions, and suffer from the limitation that depth resolution is limited by the distance separating the multiple cameras. Further, the need to provide multiple cameras adds to the cost of the overall system.

What is needed is a virtual simulation system in which a user can view and manipulate computer-generated objects and thereby control actual objects, preferably without requiring the user to wear sensor-implemented devices. Further, such system should permit other persons to see the virtual objects that are being manipulated. Such system should not require multiple image acquiring cameras (or equivalent) and should function in various lighting environments and should not be subject to inaccuracy due to changing ambient light and/or contrast. Such system should use Z-values (distance vector measurements) rather than luminosity data to recognize user interaction with system-created virtual images.

The present invention provides such a system.

#### SUMMARY OF THE INVENTION

The present invention provides computer simulations in which user-interaction with computer-generated images of objects to be manipulated is captured in three-dimensions, without requiring the user to wear sensors. The images may be projected using conventional methods including liquid crystal displays and micro-mirrors.

A computer system renders objects that preferably are viewed preferably in a heads-up display (HUD). Although neither goggles nor special viewing equipment is required by the user in an HUD embodiment, in other applications the display may indeed include goggles, a monitor, or other display equipment.

In a motor vehicle application, the HUD might be a rendering of a device for the car, e.g., a car radio, that is visible by the vehicle driver looking toward the vehicle windshield. To turn the virtual radio on, the driver would move a hand close as if to "touch" or otherwise manipulate the projected image of an on/off switch in the image. To change volume, the driver would "move" the projected image of a volume control. There is substantially instant feedback between the parameter change in the actual device, e.g., loudness of the radio audio, as perceived (e.g., heard) by the user, and user "movement" of the virtual control.

To change stations, the driver would "press" the projected image of a frequency control until the desired station is heard, whereupon the virtual control would be released by the user. Other displayed images may include warning messages concerning the state of the vehicle, or other environment, or GPS-type map  
5 displays that the user can control.

The physical location and movement of the driver's fingers in interacting with the computer-generated images in the HUD is determined non-haptically in three-dimensions by a three-dimensional range finder within the system. The three-  
10 dimensional data acquisition system operates preferably by transmitting light signals, e.g., energy in the form of laser pulses, modulated light beams, etc. In a preferred embodiment, return time-of-flight measurements between transmitted energy and energy reflected or returned from an object can provide (x,y,z) axis position information as to the presence and movement of objects. Such objects  
15 can include a user's hand, fingers, perhaps a held baton, in a sense-vicinity to virtual objects that are projected by the system. In an HUD application, such virtual objects may be projected to appear on (or behind or in front of) a vehicle windshield. Preferably ambient light is not relied upon in obtaining the three-dimensional position information, with the result that the system does not lose  
20 positional accuracy in the presence of changing light or contrast environments. In other applications, modulated light beams could instead be used.

When the user's hand (or other object evidencing user-intent) is within a sense-frustum range of the projected object, the three-dimensional range output data  
25 is used to change the computer-created image in accordance with the user's hand or finger (or other) movement. If the user hand or finger (or other) motion "moves" a virtual sliding radio volume control to the right within the HUD, the system will cause the virtual image of the slider to be moved to the right. At the same time, the volume on the actual radio in the vehicle will increase, or  
30 whatever device parameter is to be thus controlled. Range finding information is collected non-haptically, e.g., the user need not actually touch anything for (x,y,z) distance sensing to result.

The HUD system can also be interactive in the sense of displaying dynamic images as required. A segment of the HUD might be motor vehicle gages, which segment is not highlighted unless the user's fingers are moved to that region. On the other hand, the system can automatically create and highlight certain  
5 images when deemed necessary by the computer, for example a flashing "low on gas" image might be projected without user request.

In other applications, a CRT or LCD display can be used to display a computer rendering of objects that may be manipulated with a user's fingers, for example  
10 a virtual thermostat to control home temperature. "Adjusting" the image of the virtual thermostat will in fact cause the heating or cooling system for the home to be readjusted. Advantageously such display(s) can be provided where convenient to users, without regard to where physical thermostats (or other  
15 controls) may actually have been installed. In a factory training application, the user may view an actual object being remotely manipulated as a function of user movement, or may view a virtual image that is manipulated as a function of user movement, which system-detected movement causes an action object to be moved.

20 The present invention may also be used to implement training systems. In its various embodiments, the present invention presents virtual images that a user can interact with to control actual devices. Onlookers may see what is occurring in that the user is not required to wear sensor-equipped clothing, helmets, gloves, or goggles.

25

Other features and advantages of the invention will appear from the following description in which the preferred embodiments have been set forth in detail, in conjunction with the accompanying drawings.

30

#### BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 a heads-up display of a user-immersible computer simulation, according to the present invention;

FIG. 2A is a generic block diagram showing a system with which the present invention may be practiced;

FIG. 2B depicts clipping planes used to detect user-proximity to virtual images  
5 displayed by the present invention;

FIGS. 3A-3C depict use of a slider-type virtual control, according to the present invention;

10 FIG. 3D depicts exemplary additional images created by the present invention;

FIGS. 3E and 3F depict use of a rotary-type virtual control, according to the present invention;

15 FIGS. 3G, 3H, and 3I depict the present invention used in a manual training type application;

FIGS. 4A and 4B depict reference frames used to recognize virtual rotation of a rotary-type virtual control, according to the present invention; and

20 FIGS. 5A and 5B depict user-zoomable virtual displays useful to control a GPS device, according to the present invention.

#### DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

25 Fig. 1 depicts a heads-up display (HUD) application of a user-immersible computer simulation system, according to the present invention. The present invention 10 is shown mounted in the dashboard or other region of a motor vehicle 20 in which there is seated a user 30. Among other functions, system 10 computer-generates and projects imagery onto or adjacent an image region 40  
30 of front windshield 50 of vehicle 20. Image projection can be carried out with conventional systems such as LCDs, or micro-mirrors. In this embodiment, user 30 can look ahead through windshield 50 while driving vehicle 20, and can also



see any image(s) that are projected into region 40 by system 10. In this embodiment, system 10 may properly be termed a heads-up display system. Also shown in Fig. 1 are the three reference x,y,z axes. As described later herein with reference to Fig. 2B, region 40 may be said to be bounded in the z-axis by clipping planes.

User 30 is shown as steering vehicle 20 with the left hand while the right hand is near or touching a point  $p1(t)$  on or before an area of windshield within a detection range of system 10. By "detection range" it is meant that system 10 can determine in three-dimensions the location of point  $p1(t)$  as a function of time  $(t)$  within a desired proximity to image region 40. Thus,  $p1(t)$  may be uniquely defined by coordinates  $p1(t) = (x1(t), y1(t), z1(t))$ . Because system 10 has three-dimensional range finding capability, it is not required that the hand of user 30 be covered with a sensor-laden glove, as in many prior art systems. Further, since system 10 knows what virtual objects (if any) are displayed in image region 40, the interaction between the user's finger and such images may be determined. Detection in the present invention occurs non-haptically, that is it is not required that the user's hand or finger or pointer actually make physical contact with a surface or indeed anything in order to obtain the (x,y,z) coordinates of the hand, finger, or pointer.

Fig. 1 depicts a device 60 having at least one actual control 70 also mounted in vehicle 20, device 60 shown being mounted in the dashboard region of the vehicle. Device 60 may be an electronic device such as a radio, CD player, telephone, a thermostat control or window control for the vehicle, etc. As will be described, system 10 can project one or more images, including an image of device 60 or at least a control 70 from device 60.

Exemplary implementations for system 10 may be found in co-pending U.S. patent application 09/401,059 filed 22 September 1999 entitled "CMOS-Compatible Three-Dimensional Image Sensor IC", in co-pending U.S. patent application 09/502,499 filed 11 February 2000 entitled "Method and Apparatus

for Creating a Virtual Data Entry Device”, and in co-pending U.S. patent application 09/727,529 filed 28 November 2000 entitled “CMOS-Compatible Three-Dimensional Image Sensor IC”. In that a detailed description of such systems may be helpful, applicants refer to and incorporate by reference each  
5 said pending U.S. patent application. The systems described in these patent applications can be implemented in a form factor sufficiently small to fit into a small portion of a vehicle dashboard, as suggested by Fig. 1 herein. Further, such systems consume low operating power and can provide real-time (x,y,z) information as to the proximity of a user’s hand or finger to a target region, e.g.,  
10 region 40 in Fig. 1. System 100, as used in the present invention, preferably collects data at a frame rate of at least ten frames per second, and preferably thirty frames per second. Resolution in the x-y plane is preferably in the 2 cm or better range, and in the z-axis is preferably in the 1 cm to 5 cm range.

15 A less suitable candidate for a multi-dimensional imaging system might be along the lines of U.S. patent no. 5,767,842 to Korth (1998) entitled “Method and Device for Optical Input of Commands or Data”. Korth proposes the use of conventional two-dimensional TV video cameras in a system to somehow recognize what portion of a virtual image is being touched by a human hand. But  
20 Korth’s method is subject to inherent ambiguities arising from his reliance upon relative luminescence data, and upon adequate source of ambient lighting. By contrast, the applicants’ referenced co-pending applications disclose a true time-of-flight three-dimensional imaging system in which neither luminescence data nor ambient light is relied upon.

25

However implemented, the present invention preferably utilizes a small form factor, preferably inexpensive imaging system that can find range distances in three dimensions, substantially in real-time, in a non-haptic fashion. Fig. 2A is an exemplary system showing the present invention in which the range finding  
30 system is similar to that disclosed in the above-referenced co-pending U.S. patent applications. Other non-haptic three-dimensional range finding systems could instead be used, however. In Fig. 2A, system 100 is a three-dimensional

range finding system that is augmented by sub-system 110, which generates and can project via an optical system 120 computer-created object images such as 130A, 130B. Such projection may be carried out with LCDs or micro-mirrors, or with other components known in the art. In the embodiment shown, the images  
5 created can appear to be projected upon the surface of windshield 50, in front of, or behind windshield 50.

The remainder of system 100 may be as disclosed in the exemplary patent applications. An array 140 of pixel detectors 150 and their individual processing  
10 circuits 160 is provided preferably on an IC 170 that includes most if not all of the remainder of the overall system. A typical size for the array might be 100x100 pixel detectors 150 and an equal number of associated processing circuits 160. An imaging light source such as a laser diode 180 emits energy via lens system 190 toward the imaging region 40. At least some of the emitted energy will be  
15 reflected from the surface of the user's hand, finger, a held baton, etc., back toward system 100, and can enter collection lens 200. Alternatively, rather than use pulses of energy, a phase-detection based ranging scheme could be employed.

20 The time interval from start of a pulse of emitted light energy from source 190 to when some of the reflected energy is returned via lens 200 to be detected by a pixel diode detector in array 140 is measured. This time-of-flight measurement can provide the vector distance to the location on the windshield, or elsewhere, from which the energy was reflected. Clearly if a human finger (or other object)  
25 is within the imaging region 40, locations of the surface of the finger may, if desired, also be detected and determined.

System 100 preferably provides computer functions and includes a microprocessor or microcontroller system 210 that preferably includes a control  
30 processor 220, a data processor 230, and an input/output processor 240. IC 170 preferably further includes memory 250 having random access memory (RAM) 260, read-only memory (ROM) 270, and memory storing routine(s) 280 used by

the present invention to calculate vector distances, user finger movement velocity and movement direction, and relationships between projected images and location of a user's finger(s). Circuit 290 provides timing, interface, and other support functions.

5

Within array 140, each preferably identical pixel detector 150 can generate data from to calculate Z distance to a point  $p1(t)$  in front of windshield 50, on the windshield surface, or behind windshield 50, or to an intervening object. In the disclosed applications, each pixel detector preferably simultaneously acquires  
10 two types of data that are used to determine Z distance: distance time delay data, and energy pulse brightness data. Delay data is the time required for energy emitted by emitter 180 to travel at the speed of light to windshield 40 or, if closer, a user's hand or finger or other object, and back to sensor array 140 to be detected. Brightness is the total amount of signal generated by detected  
15 pulses as received by the sensor array. It will be appreciated that range finding data is obtained without touching the user's hand or finger with anything, e.g., the data is obtained non-haptically.

As shown in Fig. 2B, region 40 may be considered to be bounded in the z-axis  
20 direction from a front clipping plane 292 and by a rear clipping plane 294. Rear clipping plane 292 may coincide with the z-axis distance from system 100 to the inner surface of windshield 50 (or other substrate in another application). The z-axis distance separating planes 292 and 294 represents the proximity range within which a user's hand or forefinger is to be detected with respect to  
25 interaction with a projected image, e.g. 130B. In Fig. 2B, the tip of the user's forefinger is shown as passing through plane 292 to "touch" image 130B, here projected to appear intermediate the two clipping planes.

In reality, clipping planes 292 and 294 will be curved and the region between  
30 these planes can be defined as an immersion frustum 296. As suggested by Fig. 2B, image 130B may be projected to appear within immersion frustum 296, or to appear behind (or outside) the windshield. If desired, the image could be made

to appear in front of the frustum. The upper and lower limits of region 40 are also bounded by frustum 296 in that when the user's hand is on the car seat or on the car roof, it is not necessary that system 100 recognize the hand position with respect to any virtual image, e.g., 130B, that may be presently displayed. It will be appreciated that the relationship shown in Fig. 2B is a very intuitive way to provide feedback in that the user sees the image of a control 130B, reaches towards and appears to manipulate the control.

Three-dimensional range data is acquired by system 100 from examination of time-of-flight information between signals emitted by emitter 180 via optional lens 190, and return signals entering optional lens 200 and detected by array 140. Since system 100 knows a priori the distance and boundaries of frustum 296 and can detect when an object such as a user's forefinger is within the spaced bounded by the frustum. Software 290 recognizes the finger or other object is detected within this range, and system 100 is essentially advised of potential user intent to interact with any displayed images. Alternatively, system 100 can display a menu of image choices when an object such as a user's finger is detected within frustum 296. (For example, in Fig. 3D, display 130D could show icons rather than buttons, one icon to bring up a cellular telephone dialing display, another icon to bring up a map display, another icon to bring up vehicle control displays, etc.)

Software 290 attempts to recognize objects (e.g., user's hand, forefinger, perhaps arm and body, head, etc.) within frustum 206, and can detect shape (e.g., perimeter) and movement (e.g., derivative of positional coordinate changes). If desired, the user may hold a passive but preferably highly reflective baton to point to regions in the virtual display. Although system 100 preferably uses time-of-flight z-distance data only, luminosity information can aid in discerning objects and object shapes and positions.

30

Software 290 could cause a display that includes virtual representations of portions of the user's body. For example if the user's left hand and forefinger are

recognized by system 100, the virtual display in region 40 could include a left hand and forefinger. If the user's left hand moved in and out or left and right, the virtual image of the hand could move similarly. Such application could be useful in a training environment, for example where the user is to pickup potentially dangerous items and manipulate them in a certain fashion. The user would view a virtual image of the item, and would also view a virtual image of his or her hand grasping the virtual object, which virtual object could then be manipulated in the virtual space in frustum 296.

10 Figs. 3A, 3B, and 3C show portion 40 of an exemplary HUD display, as used by the embodiment of Fig. 1 in which system 100 projected image 130A is a slider control, perhaps a representation or token for an actual volume control 80 on an actual radio 70 within vehicle 20. As the virtual slider bar 300 is "moved" to the right, it is the function of the present invention to command the volume of radio  
15 70 to increased, or if image 130A is a thermostat, to command the temperature within vehicle 20 to change, etc. Also depicted in Fig. 3A is a system 100 projected image of a rotary knob type control 130B having a finger indent region 310.

20 In Fig. 3A, optionally none of the projected images is highlighted in that the user's hand is not sufficiently close to region 40 to be sensed by system 100. Note, however, in Fig. 3B that the user's forefinger 320 has been moved towards windshield 50 (as depicted in Fig. 1), and indeed is within sense region 40. Further, the (x,y,z) coordinates of at least a portion of forefinger 320 are  
25 sufficiently close to the virtual slider bar 300 to cause the virtual slider bar and the virtual slider control image 130A to be highlighted by system 100. For example, the image may turn red as the user's foregoing "touches" the virtual slider bar. It is understood that the vector relationship in three-dimensions between the user's forefinger and region 40 is determined substantially in real-  
30 time by system 100, or by any other system able to reliably calculate distance coordinates in three-axes. In Fig. 3B the slider bar image has been "moved" to the right, e.g., as the user's forefinger moves left to right on the windshield,

system 100 calculates the forefinger position, calculates that the forefinger is sufficiently close to the slider bar position to move the slider bar, and projects a revised image into region 40, wherein the slider bar has followed the user's forefinger.

5

At the same time, electrical bus lead 330 (see Fig. 2A), which is coupled to control systems in vehicle 20 including all devices 70 that are desired to at least have the ability to be virtually controlled, according to the present invention. Since system 100 is projecting an image associated, for example, with radio 70, the volume in radio 70 will be increased as the user's forefinger slides the computer rendered image of the slider bar to the right. Of course if the virtual control image 130 were say bass or treble, then bus lead 330 would command radio 70 to adjust bass or treble accordingly. Once the virtual slider bar image 300 has been "moved" to a desirable location by the user's forefinger, system 100 will store that location and continue to project, as desired by the user or as pre-programmed, that location for the slider bar image. Since the projected images can vary, it is understood that upon re-displaying slider control 130A at a later time (e.g., perhaps seconds or minutes or hours later), the slider bar will be shown at the last user-adjusted position, and the actual control function in device 70 will be set to the same actual level of control.

Turning to Fig. 3D, assume that no images are presently active in region 40, e.g., the user is not or has not recently moved his hand or forefinger into region 40. But assume that system 100, which is coupled to various control systems and sensors via bus lead 330, now realizes that the gas tank is nearly empty, or that tire pressure is low, or that oil temperature is high. System 100 can now automatically project an alert or warning image 130C, e.g., "ALERT" or perhaps "LOW TIRE PRESSURE", etc. As such, it will be appreciated that what is displayed in region 40 by system 100 can be both dynamic and interactive.

30

Fig. 3D also depicts another HUD display, a virtual telephone dialing pad 130D, whose virtual keys the user may "press" with a forefinger. In this instance, device

70 may be a cellular telephone coupled via bus lead 130 to system 100. As the user's forefinger touches a virtual key, the actual telephone 70 can be dialed. Software, e.g., routine(s) 280, within system 100 knows a priori the location of each virtual key in the display pad 130D, and it is a straightforward task to  
5 discern when an object, e.g., a user's forefinger, is in close proximity to region 40, and to any (x,y,z) location therein. When a forefinger hovers over a virtual key for longer than a predetermined time, perhaps 100 ms, the key may be considered as having been "pressed". The "hovering" aspect may be determined, for example, by examining the first derivative of the (x(t),y(t),z(t))  
10 coordinates of the forefinger. When this derivative is zero, the user's forefinger has no velocity and indeed is contacting the windshield and can be moved no further in the z-axis. Other techniques may instead be used to determine location of a user's forefinger (or other hand portion), or a pointer held by the user, relative to locations within region 40.

15

Referring to Fig. 3E, assume that the user wants to "rotate" virtual knob 130B, perhaps to change frequency on a radio, to adjust the driver's seat position, to zoom in or zoom out on a projected image of a road map, etc. Virtual knob 130B may be "grasped" by the user's hand, using for example the right thumb 321, the  
20 right forefinger 320, and the right middle finger 322, as shown in Fig. 3E. By "grasped" it is meant that the user simply reaches for the computer-rendered and projected image of knob 130B as though it were a real knob. In a preferred embodiment, virtual knob 130B is rendered in a highlight color (e.g., as shown by Fig. 3E) when the user's hand (or other object) is sufficiently close to the area  
25 of region 40 defined by knob 130B. Thus in Fig. 3A, knob 130B might be rendered in a pale color, since no object is in close proximity to that portion of the windshield. But in Fig. 3E, software 280 recognizes from acquired three-dimensional range finding data that an object (e.g., a forefinger) is close to the area of region 40 defined by virtual knob 130B. Accordingly in Fig. 3E, knob  
30 130B is rendered in a more discernable color and/or with bolder lines than is depicted in Fig. 3A.



In Fig. 3E, the three fingers noted will "contact" virtual knob 130B at three points, denoted a1 (thumb tip position), a2 (forefinger tip position), and a3 (middle fingertip position). With reference to Figs. 4A and 4B, analysis can be carried out by software 280 to recognize the rotation of virtual knob 130B that is shown in  
 5 Fig. 3F, to recognize the magnitude of the rotation, and to translate such data into commands coupled via bus 330 to actual device(s) 70.

Consider the problem of determining the rotation angle  $\Theta$  of virtual knob 130B given coordinates for three points a1, a2, and a3, representing perceived tips of  
 10 user fingers before rotation. System 100 can compute and/or approximate the rotation angle  $\Theta$  using any of several approaches. In a first approach, the exact rotation angle  $\Theta$  is determined as follows. Let the pre-rotation (e.g., Fig. 3E position) points be denoted  $a_1 = (x_1, y_1, z_1)$ ,  $a_2 = (x_2, y_2, z_2)$ , and  $a_3 = (x_3, y_3, z_3)$  and let  $A_1 = (X_1, Y_1, Z_1)$ ,  $A_2 = (X_2, Y_2, Z_2)$ , and  $A_3 = (X_3, Y_3, Z_3)$  be the respective  
 15 coordinates after rotation through angle  $\theta$ , as shown in Fig. 3F. In Figs. 3E and 3F and 4A and 4B, rotation of the virtual knob is shown in a counter-clockwise direction.

Referring to Fig. 4A, the center of rotation may be considered to be point  
 20  $p = (x_p, y_p, z_p)$ , whose coordinates are unknown. The axis of rotation is approximately normal to the plane of the triangle defined by the three fingertip contact points  $a_1, a_2$  and  $a_3$ . The (x,y,z) coordinates of point  $p$  can be calculated by the following formula:

$$25 \quad \begin{bmatrix} x_p \\ y_p \\ z_p \end{bmatrix} = \frac{1}{2} \begin{bmatrix} X_1 - x_1 & Y_1 - y_1 & Z_1 - z_1 \\ X_2 - x_2 & Y_2 - y_2 & Z_2 - z_2 \\ X_3 - x_3 & Y_3 - y_3 & Z_3 - z_3 \end{bmatrix}^{-1} \begin{bmatrix} X_1^2 + Y_1^2 + Z_1^2 - x_1^2 - y_1^2 - z_1^2 \\ X_2^2 + Y_2^2 + Z_2^2 - x_2^2 - y_2^2 - z_2^2 \\ X_3^2 + Y_3^2 + Z_3^2 - x_3^2 - y_3^2 - z_3^2 \end{bmatrix}.$$

If the rotation angle  $\theta$  is relatively small, angle  $\theta$  can be calculated as follows:  
 30

$$\theta = \sqrt{\frac{X_i^2 + Y_i^2 + Z_i^2 - x_i^2 - y_i^2 - z_i^2}{(x_i - x_p)^2 + (y_i - y_p)^2 + (z_i - z_p)^2}} \text{ for } i=1,2, \text{ or } 3.$$

Alternatively, system 100 may approximate rotation angle  $\Theta$  using a second  
 5 approach, in which an exact solution is not required. In this second approach,  
 it is desired to ascertain direction of rotation (clockwise or counter-clockwise)  
 and to approximate the magnitude of the rotation.

Referring now to Fig. 4C, assume that point  $c = (c_x, c_y, c_z)$  is the center of the  
 10 triangle defined by the three pre-rotation points  $a_1, a_2$  and  $a_3$ . The following  
 formula may now be used:

$$\begin{cases} c_x = \frac{x_1 + x_2 + x_3}{3} \\ c_y = \frac{y_1 + y_2 + y_3}{3} \\ c_z = \frac{z_1 + z_2 + z_3}{3} \end{cases}$$

Again, as shown in Fig. 1, the z-axis extends from system 100, and the x-axis  
 15 and y-axis are on the plane of the array of pixel diode detectors 140. Let L  
 be a line passing through points  $a_1, a_2$ , and let  $L_{xy}$  be the projection of line L  
 onto the x-y plane. Line  $L_{xy}$  may be represented by the following equation:

$$L(x, y) \equiv \frac{y_1 - y_2}{x_2 - x_1} (x - x_1) + y - y_1 = 0$$

The clockwise or counter-clockwise direction of rotation may be defined by the following criterion:

Rotation is clockwise if  $L(c_x, c_y) \cdot L(X_2, Y_2) < 0$ , and rotation is counter-clockwise if  $L(c_x, c_y) \cdot L(X_2, Y_2) > 0$ .

5

When  $L(c_x, c_y) \cdot L(X_2, Y_2) = 0$ , a software algorithm, perhaps part of routine(s)

290, executed by computer sub-system 210 selects points  $a_2, a_3$ , passes line

L through points  $a_2, a_3$ , and uses the above criterion to define the direction of rotation. The magnitude of rotation may be approximated by defining  $d_i$ , the

10 distance between  $a_i$  and  $A_i$ , as follows:

$$d_i = \sqrt{(X_i - x_i)^2 + (Y_i - y_i)^2 + (Z_i - z_i)^2} \text{ for } i=1,2,3.$$

The magnitude of the rotation angle  $\Theta$  may be approximated as follows:

$$\theta \cong k(d_1 + d_2 + d_3),$$

15 where  $k$  is a system constant that can be adjusted.

The analysis described above is somewhat generalized to enable remote tracking of rotation of any three points. A more simplified approach may be used in Fig. 3E, where user 30 may use a fingertip to point to virtual

20 indentation 310 in the image of circular knob 130B. The fingertip may now move clockwise or counter-clockwise about the rotation axis of knob 130B, with the result that system 100 causes the image of knob 130B to be rotated to track the user's perceived intended movement of the knob. At the same time, an actual controlled parameter on device 70 (or vehicle 20) is moved,  
25 proportionally to the user movement of the knob image. As in the other

embodiments, the relationship between user manipulation of a virtual control and variation in an actual parameter of an actual device may be linear or otherwise, including linear in some regions of control and intentionally non-linear in other regions.

5

Software 290 may of course use alternative algorithms, executed by computer system 210, to determine angular rotation of virtual knobs or other images rendered by computing system 210 and projected via lens 190 onto windshield or other area 50. As noted, computing system 210 will then generate the  
10 appropriate commands, coupled via bus 330 to device(s) 70 and/or vehicle 20.

Figs. 3G and 3H depict use of the present invention as a virtual training tool in which a portion of the user's body is immersed in the virtual display. In this application, the virtual display 40' may be presented on a conventional monitor  
15 rather than in an HUD fashion. As such, system 100 can output video data and video drive data to a monitor, using techniques well known in the art. For ease of illustration, a simple task is shown. Suppose the user, whose hand is depicted as 302, is to be trained to pick up an object, whose virtual image is shown as 130H (for example a small test tube containing a highly dangerous substance),  
20 and to carefully tilt the object so that its contents pour out into a target region, e.g., a virtual beaker 130I. In Fig. 3G, the user's hand, which is detected and imaged by system 100, is depicted as 130G in the virtual display. For ease of illustration, virtual hand 130G is shown as a stick figure, but a more realistic image may be rendered by system 100. In Fig. 3H, the user's real hand 302 has  
25 rotated slightly counter-clockwise, and the virtual image 40' shows virtual object 130H and virtual hand 130G similarly rotated slightly counter-clockwise.

The sequence can be continued such that the user must "pour out" virtual contents of object 130H into the target object 130I without spilling. System 100  
30 can analyze movements of the actual hand 302 to determine whether such movements were sufficiently carefully executed. The virtual display could of course depict the pouring-out of contents, and if the accuracy of the pouring were

not proper, the spilling of contents. Object 130H and/or its contents (not shown) might, for example, be highly radioactive, and the user's hand motions might be practice to operate a robotic control that will grasp and tilt an actual object whose virtual representation is shown as 130H. However use of the present invention  
5 permits practice sessions without the risk of any danger to the user. If the user "spills" the dangerous contents or "drops" the held object, there is no harm, unlike a practice session with an actual object and actual contents.

Fig. 3I depicts the present invention used in another training environment. In this  
10 example, user 302 perhaps actually holds a tool 400 to be used in conjunction with a second tool 410. In reality the user is being trained to manipulate a tool 400' to be used in conjunction with a second tool 410', where tool 400' is manipulated by a robotic system 420, 430 (analogous to device 70) under control of system 100, responsive to user-manipulation of tool 400. Robotically  
15 manipulated tools 400', 410' are shown behind a pane 440, that may be a protective pane of glass, or that may be opaque, to indicate that tools 400', 410' cannot be directly viewed by the user. For example, tools 400', 410' may be at the bottom of the ocean, or on the moon, in which case communication bus 330 would include radio command signals. If the user can indeed view tools 400',  
20 410' through pane 440, there would be no need for a computer-generated display. However if tools 400', 410' cannot be directly viewed, then a computer-generated display 40' could be presented. In this display, 130G could now represent the robotic arm 420 holding actual tool 400'. It is understood that as the user 302 manipulates tool 400 (although manipulation could occur without  
25 tool 400), system 100 via bus 330 causes tool 400' to be manipulated robotically. Feedback to the user can occur visually, either directly through pane 440 or via display 40', or in terms of instrumentation that in substantial real-time tells the user what is occurring with tools 400, 410'.

30 Thus, a variety of devices 70 may be controlled with system 100. Fig. 5A depicts a HUD virtual display created and projected by system 100 upon region 40 of windshield 50, in which system 70 is a global position satellite (GPS) system, or

perhaps a computer storing zoomable maps. In Fig. 5A, image 130E is shown as a roadmap having a certain resolution. A virtual scroll-type control 130F is presented to the right of image 130E, and a virtual image zoom control 130A is also shown. Scroll control 130F is such that a user's finger can touch a portion of the virtual knob, e.g., perhaps a north-east portion, to cause projected image 130E to be scrolled in that compass direction. Zoom control 130A, shown here as a slider bar, permits the user to zoom the image in or out using a finger to "move" virtual slider bar 300. If desired, zoom control 130A could of course be implemented as a rotary knob or other device, capable of user manipulation.

10

In Fig. 5B, the user has already touched and "moved" virtual slider bar 300 to the right, which as shown by the indicated portion of image 130A has zoomed in image 130E. Thus, the image, now denoted 130E, has greater resolution and provides more details. As system 100 detects the user's finger (or pointer or other object) near bar 300, detected three-dimensional (x,y,z) data permits knowing what level of zoom is desired. System 100 then outputs on bus 330 the necessary commands to cause GPS or computer system 70 to provide a higher resolution map image. Because system 100 can respond substantially in real-time, there is little perceived lag between the time the user's finger "slides" bar 300 left or right and the time map image 130E is zoomed in or out. This feedback enables the user to rapidly cause the desired display to appear on windshield 50, without requiring the user to divert attention from the task of driving vehicle 20, including looking ahead, right through the images displayed in region 40, to the road and traffic ahead.

25

Modifications and variations may be made to the disclosed embodiments without departing from the subject and spirit of the invention as defined by the following claims.

30

WHAT IS CLAIMED IS:

1. A method of presenting a virtual simulation to control an actual  
5 device, the method comprising the following steps:
  - (a) generating a display including an image of a control to change a parameter of said device;
  - (b) sensing (x,y,z) axes proximity of a user to said image on said display;
  - 10 (c) determining non-haptically from data sensed at step (b), user intended movement of said image of said control; and
  - (d) outputting a signal coupleable to said actual device to control said parameter as a function of sensed user intended movement of said image of said control.
- 15 2. The method of claim 1, wherein at step (a), said display is a heads-up-display.
3. The method of claim 1, wherein step (b) includes sensing using  
20 time-of-flight data.
4. The method of claim 1, wherein step (c) includes modifying said display to represent movement of said control created by said user.
- 25 5. The method of claim 1, wherein step (a) includes generating an image of a slider control.
6. The method of claim 1, wherein step (a) includes generating an image of a rotary control.
- 30 7. The method of claim 1, wherein step (a) includes generating an image including a menu of icons selectable by said user.

8. The method of claim 1, wherein said actual device is selected from a group consisting of (i) an electronic entertainment device, (ii) radio, (iii) a cellular telephone, (iv) a heater system, (v) a cooling system, (vi) a motorized system.

5

9. The method of claim 1, wherein at step (a) said display is generated only after detection of a user in close proximity to an area whereon said display is presentable.

10 10. The method of claim 9, further including displaying a user-alert warning responsive to a parameter of said device, independently of user proximity to said area.

11. The method of claim 1, wherein said display is a heads-up-display  
15 in a motor vehicle operable by a user, and said device is selected from a group consisting of (i) said motor vehicle, and (ii) an electronic accessory disposed in said motor vehicle.

12. The method of claim 11, wherein said device is a global position  
20 satellite system, said display includes a map, and said control is user-operable to change displayed appearance of said map.

13. A method of presenting a virtual simulation, the method comprising the following steps:

- 25 (a) generating a display including a virtual image of an object;  
(b) non-haptically sensing in three-dimensions proximity of at least a portion of a user's body to said display;  
(c) modifying said display substantially in real-time to include a representation of said user's body; and  
30 (d) modifying said display to depict substantially in real-time said representation of said user's body manipulating said object.



14. The method of claim 13, wherein said manipulating is part of a regime to train said user to manipulate a real object represented by said virtual image.

5 15. A virtual simulation system, comprising:  
an imaging sub-system to generate a display including an image;  
a detection sub-system to non-haptically detect in three-dimensions  
proximity of a portion of an object to a region of said display; and  
said imaging sub-system modifying said image in response to detected  
10 proximity of said portion of said object.

16. The system of claim 15, wherein said image is a representation of a control, said object is a portion of a user's hand, and said proximity includes user manipulation of said image; further including:  
15 a system outputting a signal coupleable to a real device having a parameter variable in response to said user manipulation of said image.

17. The system of claim 15, wherein:  
said system is a heads-up-system;  
20 said display is presentable on a windshield of a motor vehicle; and  
said image includes an image of a control.

18. The system of claim 17, wherein:  
said system includes a circuit outputting a command signal responsive to  
25 said detection of said proximity, said command signal coupleable to a device selected from a group consisting of (a) an electrically-controllable component of said motor vehicle, (b) an electrically-controllable electronic device disposed in said motor vehicle.

30 19. The system of claim 18, wherein said device is a global positioning satellite (GPS) system, wherein said image is a map generated by said GPS system, and said image is a control to change appearance of said image of said

map.

20. The system of claim 17, wherein said detection sub-system operates independently of ambient light.

5

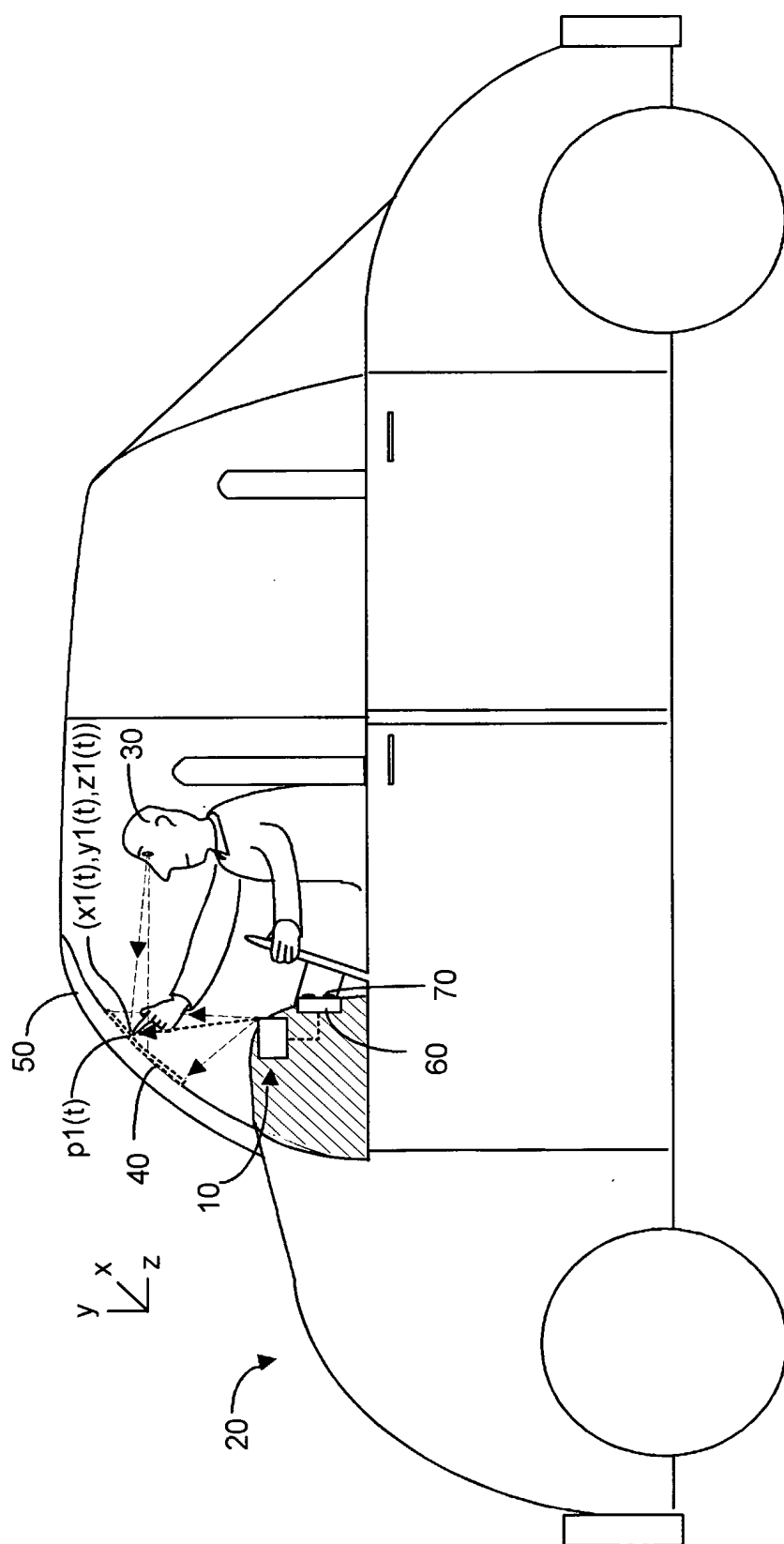


FIG. 1

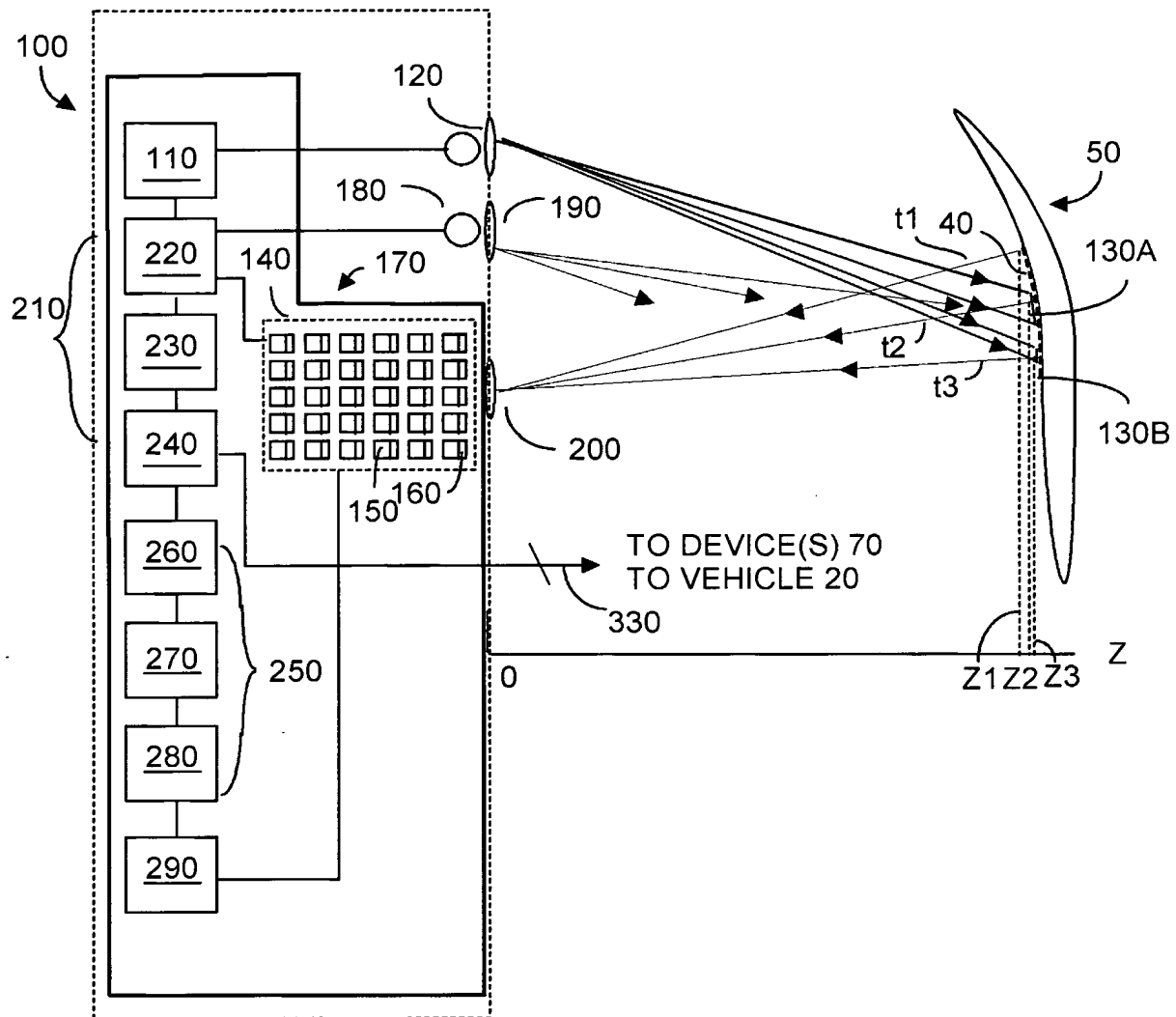


FIG. 2A

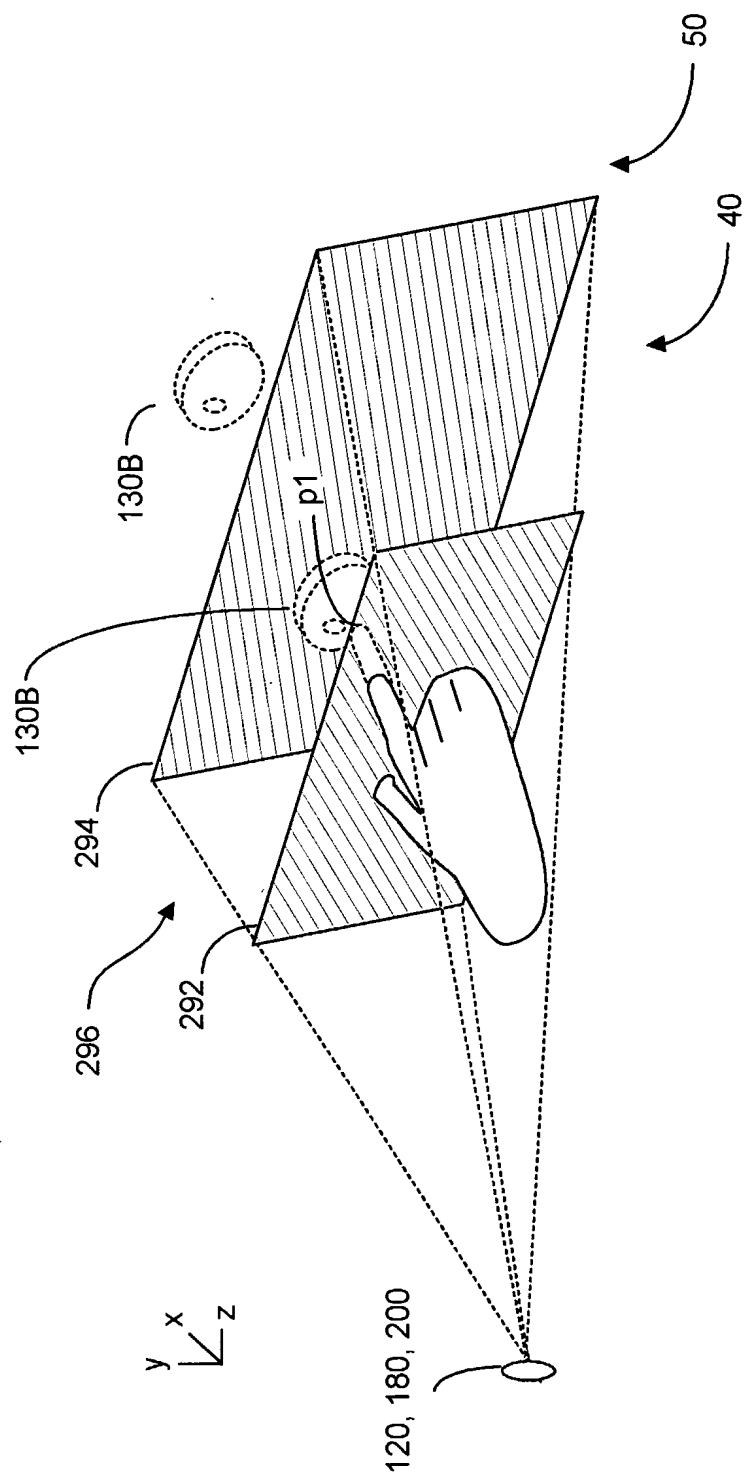


FIG. 2B

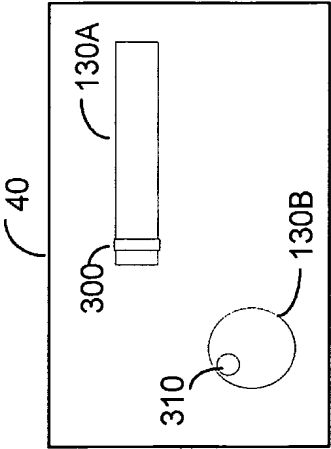


FIG. 3A

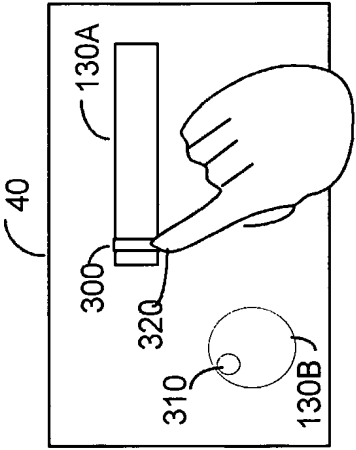


FIG. 3B

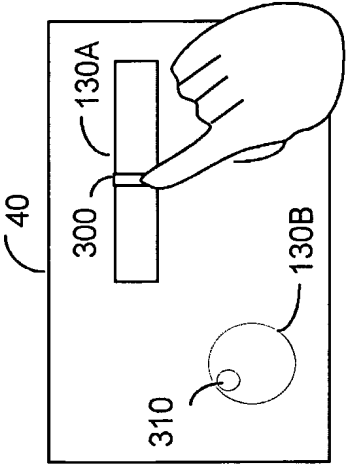


FIG. 3C

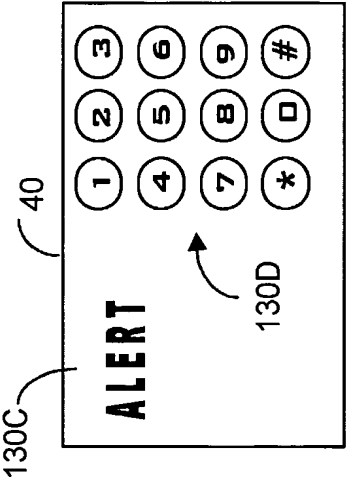


FIG. 3D

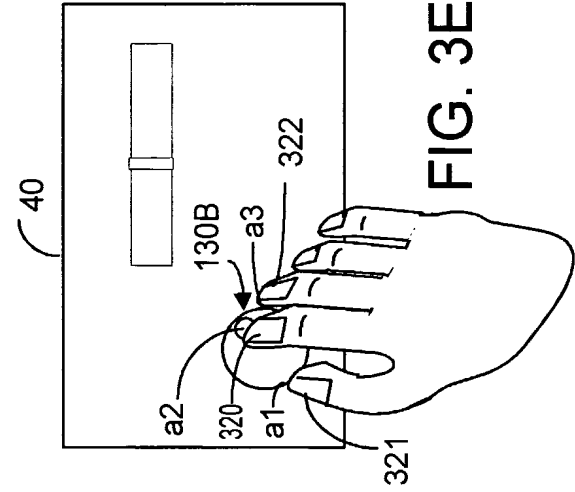


FIG. 3E

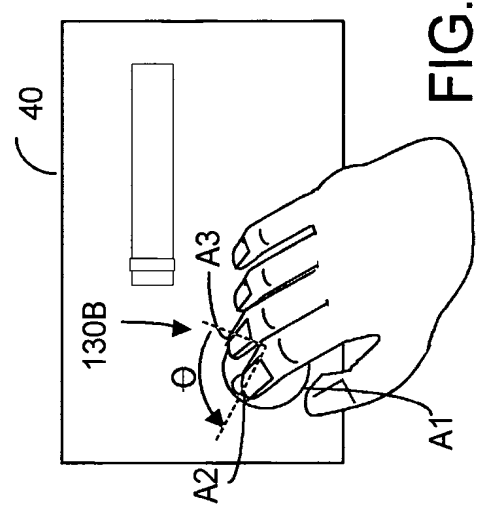


FIG. 3F

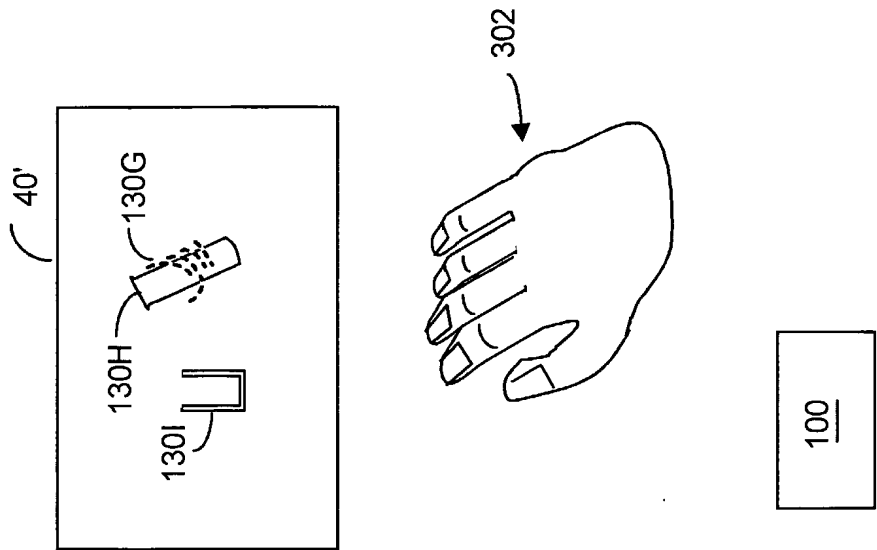


FIG. 3H

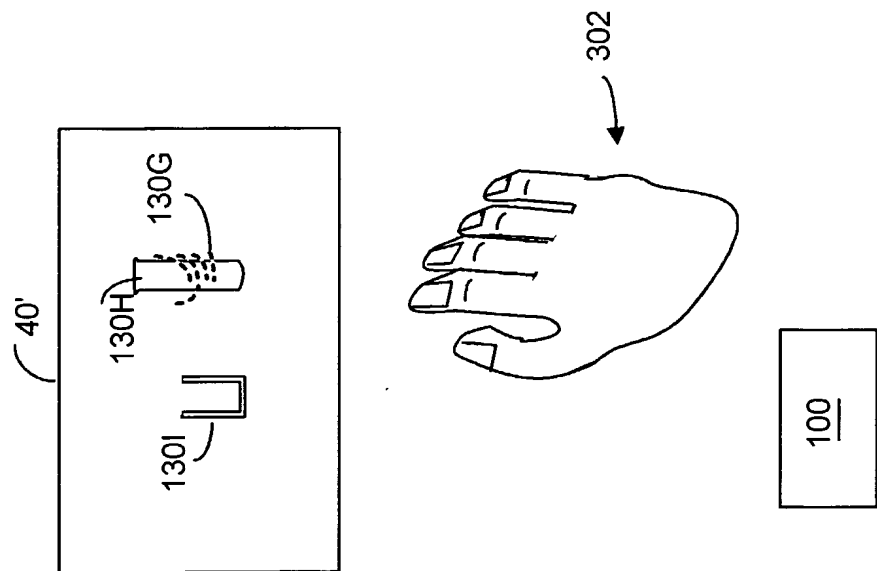


FIG. 3G

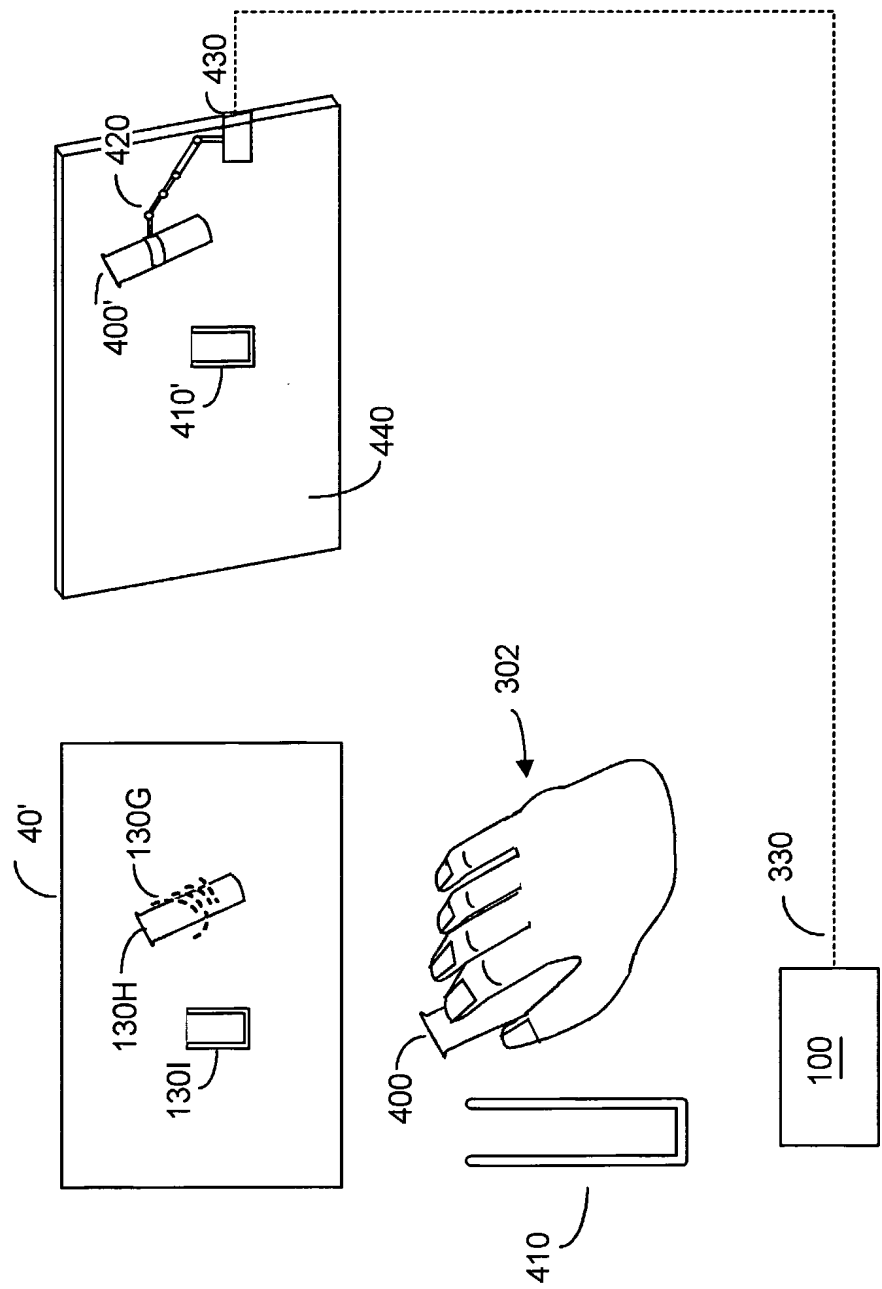


FIG. 3I



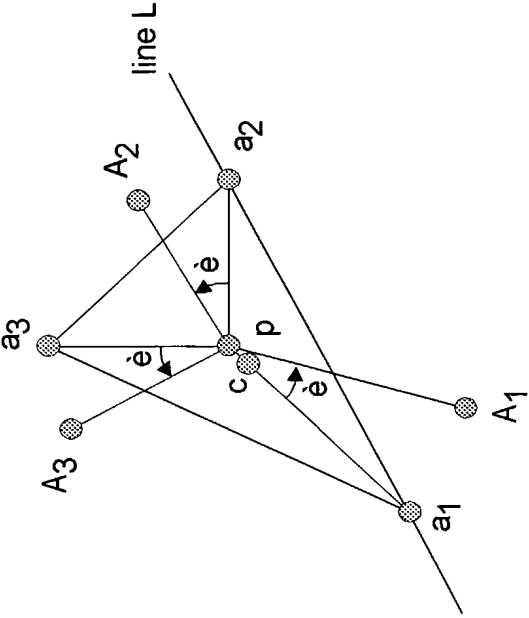


FIG. 4B

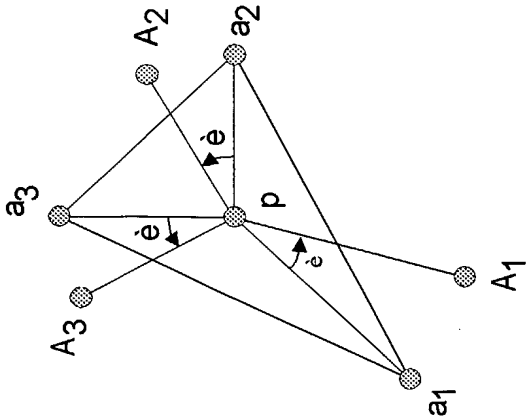


FIG. 4A

FIG. 5A

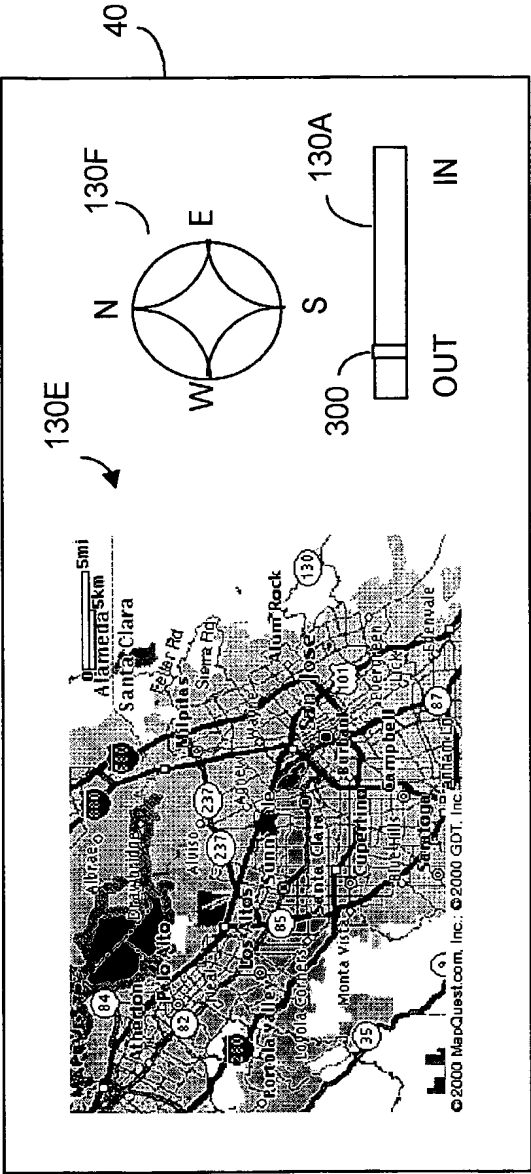
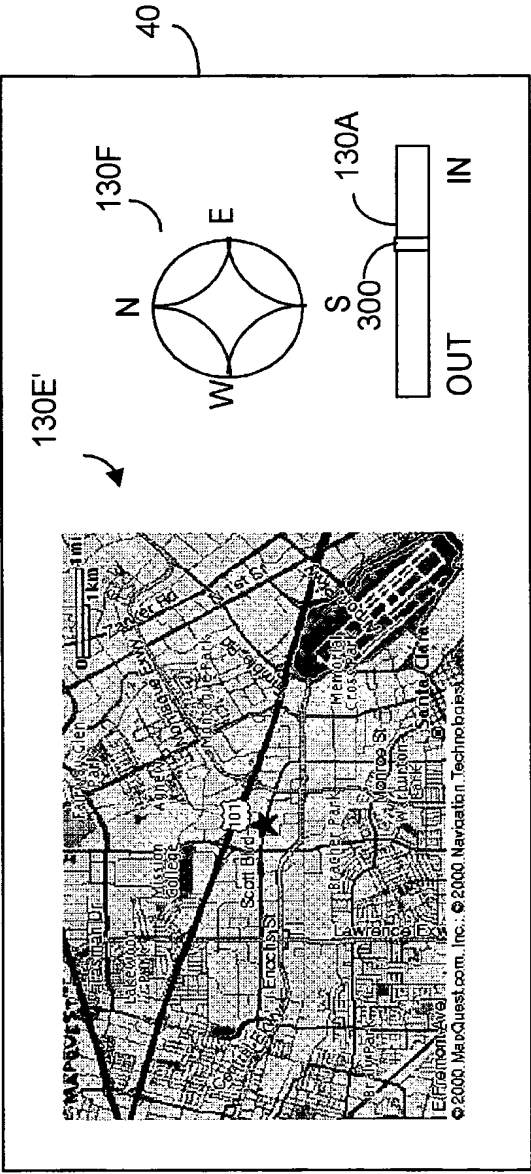


FIG. 5B



# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/03433

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : G09G 5/00, 5/08; G03B 21/00

US CL : 345/7, 8, 158, 773, 863; 353/11, 12, 13, 14

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 345/7, 8, 158, 773, 863; 353/11, 12, 13, 14

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 4,818,048 A (MOSS) 04 April 1989 (4.04.1989), column 1, lines 27-54, column 2, lines 32-68; column 3, lines 3-15; column 5, lines 19-40.	1-20
Y	US 5,990,865 A (GARD) 23 November 1999 (23.11.1999), column 1, lines 64-67; column 2, lines 1-26, 62-67; column 3, lines 1-12; column 5, lines 53-62; abstract.	1-20
Y	US 6,115,128 A (VANN) 05 September 2000 (5.09.2000), abstract	3
Y	US 5,721,679 A (MONSON) 24 February 1998 (24.02.1998), column 5, lines 58-67; column 6, lines 49-59; column 7, lines 1-9	12, 19
Y	US 5,831,584 A (SOCKS et al) 03 November 1998 (3.11.1998), column 4, lines 8-50; abstract	14
Y	US 5,486,840 A (BORREGO et al) 23 January 1996 (23.01.1996), abstract	20

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

\* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

03 June 2002 (03.06.2002)

Date of mailing of the international search report

03 JUL 2002

Name and mailing address of the ISA/US

Commissioner of Patents and Trademarks

Box PCT

Washington, D.C. 20231

Facsimile No. (703)305-3230

Authorized officer

Michael J Moyer

Telephone No. (703) 305-2099